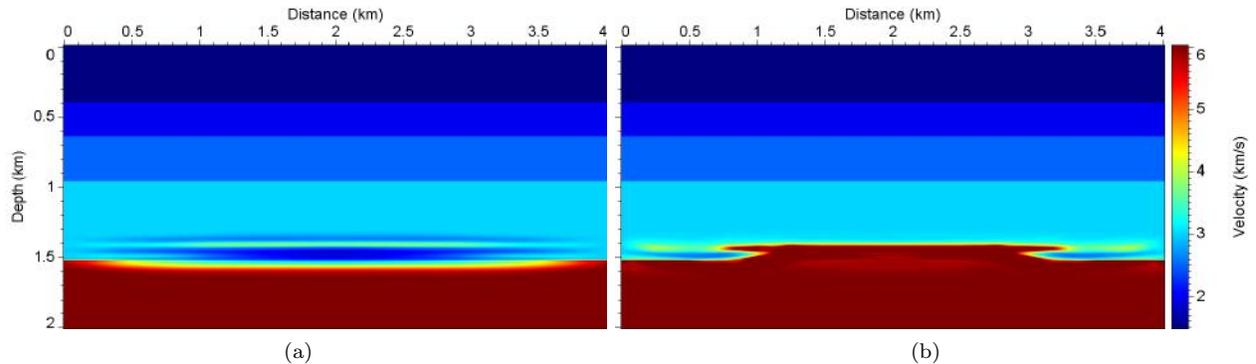# Full waveform inversion with dynamic image warping

Yong Ma & Dave Hale

Center for Wave Phenomena, Colorado School of Mines, Golden, CO 80401, USA

Figure 1. Velocities estimated by conventional reflection FWI (a) and the hybrid FWI with dynamic image warping (b). The correct depth of the deep reflector is at 1.4 km.

ABSTRACT

Full waveform inversion (FWI) can generate high-resolution subsurface models, but often suffers from an objective function with local minima caused mainly by an absence of low frequencies in seismograms. These local minima cause cycle skipping when the initial model for FWI is far from the true model. To avoid cycle skipping, traveltime inversion is often used to compute initial models for FWI.

We propose to incorporate the merits of traveltime inversion in FWI. We use dynamic image warping (DIW) to measure the traveltime misfit between recorded data and synthetic data. When compared with correlation-based techniques often used in traveltime inversion, DIW reduces errors in estimates of time shifts caused by cycle skipping. FWI with DIW then uses these time shifts to mitigate the problem of local minima in the objective function. FWI with DIW is then a hybrid method for inversion that combines the benefits of both conventional FWI and traveltime inversion.

Key words: full waveform inversion, traveltime inversion, dynamic image warping

1 INTRODUCTION

Full waveform inversion (FWI) (Tarantola, 1984; Pratt, 1999) and reflection traveltime tomography (Stork, 1992; Woodward, 1992; Vasco and Majer, 1993; Zelt and Barton, 1998) are two types of methods that geophysicists use to estimate subsurface parameters, such as seismic velocity models. Both have their own advantages and disadvantages.

FWI uses recorded seismic data $\mathbf{d}_0\left(\mathbf{x}, t\right)$ to estimate parameters of a subsurface model $\mathbf{m}$, by minimizing the difference between recorded data $\mathbf{d}_0$ and synthetic data $\mathbf{d}\left(\mathbf{x}, t\right) \equiv \mathcal{F}\left(\mathbf{m}\right)$, where $\mathcal{F}$ is a forward operator that synthesizes data. FWI is usually formulated as an optimization problem (Tarantola, 2005), in which we consider the minimization of a multivariate nonlinear objective function $E\left(\mathbf{m}\right)$. The objective function can have various forms (Crase et al., 1990), e.g., L2 norm (Taran-

tola, 1984), L1 norm (Loris et al., 2007), and Huber norm (Guitton and Symes, 2003). In this paper, we use a least-squares form:

$$E\left(\mathbf{m}\right) = \frac{1}{2}\|\mathbf{d}\left(\mathbf{x}, t\right) - \mathbf{d}_0\left(\mathbf{x}, t\right)\|^2 , \qquad (1)$$

where $\|.\|$ denotes an L2 norm.

In principle, all information in recorded seismic waveforms should be taken into account in minimizing this difference in FWI. FWI comprehensively minimizes differences in traveltimes, amplitudes, converted waves, multiples, etc. between recorded and synthetic data. This all-or-nothing approach distinguishes FWI from other methods, such as traveltime tomography that focuses on only traveltime differences. Because FWI uses all of the information in the data, it can update the model with high-resolution details. However, FWI often suffers from problems such as local minima and cycle skipping (Snieder et al., 1989). These problems often arise when the initial model is poor or the recorded data do not have sufficient low frequencies.

Ray-based or wave-equation traveltime tomography uses only traveltime differences $\Delta\tau$ between recorded data $\mathbf{d}_0$ and synthetic data $\mathbf{d}$. Traveltime tomography can also be solved as a least squares optimization problem, where we minimize the following objective function:

$$E\left(\mathbf{m}\right) = \frac{1}{2}\|\Delta\tau\|^2 . \qquad (2)$$

As noted in Luo and Schuster (1991), traveltime tomography is constrained by a high-frequency assumption about the data. This assumption causes traveltime tomography to fail to estimate velocity variations with wavelengths similar to or smaller than those in the source wavelet. As a result, compared to FWI, traveltime tomography only updates a velocity model with large-scale variations. Nevertheless, large-scale model update is important for FWI. When recorded data lack low frequencies, FWI cannot estimate large-scale components of a velocity model and can possibly lead inversion to local minima. Therefore, it is a common practice to use traveltime inversion to compute initial models for FWI.

Common methods for measuring the traveltime difference $\Delta\tau$ are based on correlations of recorded data $\mathbf{d}_0$ and synthetic data $\mathbf{d}$ (Luo and Schuster, 1991; van Leeuwen and Mulder, 2010). Unfortunately, correlation-based methods also suffer from the cycle-skipping problem. In this paper, we use dynamic image warping (DIW) (Hale, 2012) to measure the traveltime misfit. DIW, as a global optimization, can avoid cycle skipping. We then design FWI with DIW by using the DIW-estimated traveltime misfit in FWI. With the traveltime misfit, we propose two methods: traveltime dominant FWI and hybrid FWI. Synthetic tests demonstrate that traveltime dominant FWI avoids cycle skipping and hybrid FWI takes the benefits of both traveltime dominant FWI and conventional FWI.
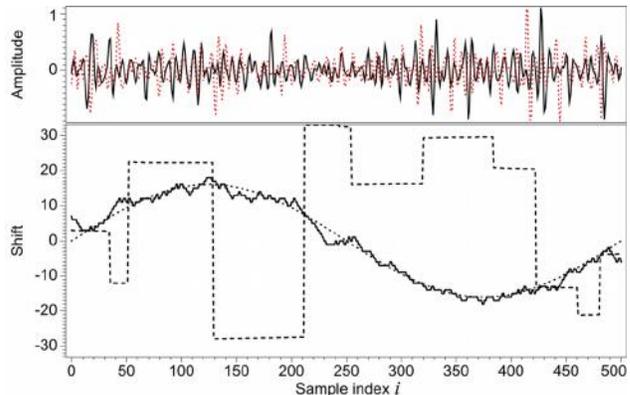
Figure 2. Top: a synthetic seismic trace (solid) and a shifted trace (dotted). Bottom: true traveltime misfit (dotted), misfit found by crosscorrelation (dashed) and dynamic warping (solid).

2 TRAVELTIME MISFIT

An essential part of traveltime tomography is estimating traveltime misfit (time shifts) between recorded and synthetic seismograms. Manually picking arrivals in seismograms is the most reliable way to do this, but is time-consuming. More automatic approaches compute crosscorrelations of seismograms to detect a traveltime misfit (Luo and Schuster, 1991; van Leeuwen and Mulder, 2010).

2.1 Crosscorrelation

Crosscorrelation of recorded and synthetic data is defined as $\int dt\ \mathbf{d}_0\left(\mathbf{x}, t\right)\mathbf{d}\left(\mathbf{x}, t+\tau\right)$. The traveltime misfit $\Delta\tau\left(\mathbf{x}\right)$ therefore is the τ that maximizes the crosscorrelation:

$$\Delta\tau\left(\mathbf{x}\right) = \arg\max_{\tau} \int dt\ \mathbf{d}_0\left(\mathbf{x}, t\right)\mathbf{d}\left(\mathbf{x}, t+\tau\right) . \qquad (3)$$

A localized traveltime misfit $\Delta\tau\left(\mathbf{x}, t\right)$ can be derived from a localized crosscorrelation:

$$\Delta\tau\left(\mathbf{x}, t\right) = \arg\max_{\tau} \int_{t-\frac{\Delta T}{2}}^{t+\frac{\Delta T}{2}} dt\ \mathbf{d}_0\left(\mathbf{x}, t\right)\mathbf{d}\left(\mathbf{x}, t+\tau\right) , \qquad (4)$$

where ΔT is the correlation window size.

Finding a maximum in a local crosscorrelation often causes cycle skipping. We illustrate this cycle-skipping problem in Figure 2 with two sequences $f[i]$ and $g[i]$. We computed the sequence $f[i]$ by convolving a Ricker wavelet with a random reflectivity sequence. We then applied time-varying shifts to that reflectivity sequence and convolved again with the same wavelet to obtain the sequence $g[i]$, which can be regarded as a shifted version of $f[i]$ plus band-limited noise. In the example of Figure 2, the shifts are a simple sinusoidal function that is plotted as the dotted line in the lower panel.

Local crosscorrelation, in this example, fails to detect the shifts; the estimated shifts (dashed line) are significantly deviated from the true shifts (dotted line) due to cycle skipping.

Another shortcoming of this correlation-based approach is that it works only if the source spectra of the recorded and the synthetic data are the same (de Hoop and van der Hilst, 2005). Therefore, van Leeuwen and Mulder (2010) modified the correlation-based approach with a weighted norm to address this shortcoming.

2.2 Dynamic image warping

To deal with the cycle-skipping problem, we choose dynamic image warping (DIW) (Hale, 2012) as a method to estimate the traveltime misfit. DIW obtains the traveltime misfit $\mathbf{u}(\mathbf{x}, t)$ between the recorded data $\mathbf{d}(\mathbf{x}, t)$ and the synthetic data $\mathbf{d}_0(\mathbf{x}, t)$ by finding a shift field $\mathbf{s}(\mathbf{x}, t)$ that solves the following constrained optimization problem:

$$\mathbf{u}(\mathbf{x}, t) = \underset{\mathbf{s}(\mathbf{x},t)}{\arg \min}\, D\left(\mathbf{s}(\mathbf{x}, t)\right) \;, \qquad (5)$$

where

$$D\left(\mathbf{s}(\mathbf{x}, t)\right) = \|\mathbf{d}(\mathbf{x}, t) - \mathbf{d}_0(\mathbf{x}, t + \mathbf{s}(\mathbf{x}, t))\|^2 \;, \qquad (6)$$

subject to the constraint

$$\left| \frac{\partial \mathbf{u}(\mathbf{x}, t)}{\partial t} \right| \leq 1 \;. \qquad (7)$$

The constraint in equation 7 is adapted from the simplest slope constraint of Sakoe and Chiba (1978). This constraint ensures that the traveltime misfit $\mathbf{u}(\mathbf{x}, t)$ neither decreases nor increases too rapidly in the t axis.

Unlike the local crosscorrelation-based approach, DIW is a global optimization approach that can avoid cycle skipping. Despite the added noise, DIW correctly detects the traveltime misfit between $f[i]$ and $g[i]$, as indicated by the solid line in the lower panel of Figure 2.

3 FWI WITH DYNAMIC IMAGE WARPING

After we obtain the traveltime misfit $\mathbf{u}(\mathbf{x}, t)$ between the recorded data $\mathbf{d}(\mathbf{x}, t)$ and the synthetic data $\mathbf{d}_0(\mathbf{x}, t)$, we can use this misfit in FWI. To do this, we need to formulate objective functions that are minimized (or maximized) when the misfit $\mathbf{u}(\mathbf{x}, t)$ is zero, indicating the reflections in the recorded data $\mathbf{d}(\mathbf{x}, t)$ and in the synthetic data $\mathbf{d}_0(\mathbf{x}, t)$ are well aligned.

3.1 FWI implementation

A typical implementation of FWI based on conjugate gradients consists of three steps performed iteratively,

beginning with an initial model $\mathbf{m}_0$. In the ith iteration, we

(i) compute the gradient $\mathbf{g}_i \equiv \mathbf{g}(\mathbf{m}_i) = \dfrac{\partial E(\mathbf{m}_i)}{\partial \mathbf{m}_i}$ with the adjoint-state method (Tromp et al., 2005);
(ii) search for a step length α_i in the conjugate gradient direction $\mathbf{h}_i\left(\mathbf{g}_i, \mathbf{g}_{i-1}\right)$ (Vigh and Starr, 2008);
(iii) update the model with $\mathbf{m}_{i+1} = \mathbf{m}_i - \alpha_i \mathbf{h}_i$;

The adjoint-state method computes the gradient as

$$\mathbf{g}_i = \mathcal{F}^*\left(\mathcal{A}(\mathbf{m}_i)\right) \;, \qquad (8)$$

where $\mathcal{F}^*$ is the adjoint of the forward operator $\mathcal{F}$, and $\mathcal{A}(\mathbf{m}_i)$ is the adjoint source. The adjoint source for FWI is defined as

$$\mathcal{A}(\mathbf{m}) = \frac{\partial E(\mathbf{m})}{\partial \mathbf{d}} \;, \qquad (9)$$

which can be rewritten as

$$\mathcal{A}(\mathbf{m}) = \mathbf{d} - \mathbf{d}_0 \;. \qquad (10)$$

Therefore, the gradient $\mathbf{g}_i$ in FWI becomes $\mathcal{F}^*(\mathbf{d} - \mathbf{d}_0)$, indicating a reverse time migration of the data residual $\mathbf{d} - \mathbf{d}_0$.

3.2 Traveltime dominant FWI

The most straightforward way of using the misfit $\mathbf{u}(\mathbf{x}, t)$ is analogous to conventional traveltime tomography, in which, we aim to solve a least squares inverse problem:

$$E(\mathbf{m}) = \frac{1}{2}\|\mathbf{u}(\mathbf{x}, t)\|^2 \;. \qquad (11)$$

If we use the three-step iterative approach outlined above to solve this optimization problem, we first must compute the gradient of the objective function in equation 11 with respect to the model parameters $\mathbf{m}$. The adjoint-state method is an efficient way to compute the gradient, and for this objective function, the adjoint source is

$$\mathcal{A}(\mathbf{m}) = \mathbf{u}\frac{\partial \mathbf{u}}{\partial \mathbf{d}} \;. \qquad (12)$$

However this adjoint source is difficult to compute because we have no close-form expression for $\partial \mathbf{u}/\partial \mathbf{d}$.

To avoid the difficulty in evaluating $\partial \mathbf{u}/\partial \mathbf{d}$, we choose another objective function that involves both the traveltime misfit $\mathbf{u}(\mathbf{x}, t)$ and the synthetic data $\mathbf{d}(\mathbf{x}, t)$:

$$E(\mathbf{m}) = \frac{1}{2}\|\mathbf{u}(\mathbf{x}, t)\,\mathbf{d}(\mathbf{x}, t)\|^2 \;. \qquad (13)$$

When the misfit $\mathbf{u}(\mathbf{x}, t)$ is zero, the objective function in equation 13 still attains a minimum. Because $\|\mathbf{d}(\mathbf{x}, t)\|^2$ cannot be zero, equation 13 is dominated by the traveltime misfit $\mathbf{u}(\mathbf{x}, t)$. We then refer to the inverse problem implied by equation 13 as traveltime dominant FWI.

The adjoint source in this case becomes

$$\mathcal{A}(\mathbf{m}) = \mathbf{u}\mathbf{d}^2\frac{\partial \mathbf{u}}{\partial \mathbf{d}} + \mathbf{u}^2\mathbf{d} \;, \qquad (14)$$

which we approximate as

$$\mathcal{A}(\mathbf{m}) \approx \mathbf{u}^2 \mathbf{d} \ . \qquad (15)$$

The assumption behind this approximation is that the traveltime misfit $\mathbf{u}$, as an implicit function of the synthetic data $\mathbf{d}$, varies slowly so that $\partial \mathbf{u}/\partial \mathbf{d}$ is small and we can discard the first term on the right hand side of equation 14. The validity of this assumption has not yet been strictly tested, and it may be wrong.

3.3 Hybrid FWI

The objective function in Equation 13 becomes zero when traveltimes in recorded and synthetic data are well aligned, even though amplitudes in synthetic data are incorrect. The traveltime dominant FWI implied by equation 13 primarily uses traveltime information, and as a consequence, solving traveltime dominant FWI implied by equation 13 is similar to performing a traveltime tomography. As discussed in the introduction, traveltime tomography recovers only the large-scale components of a velocity model (Luo and Schuster, 1991). On the other hand, FWI is not limited by traveltime and its objective function uses all the information in the data. Therefore, FWI can recover a velocity model with fine details.

To gain the advantages of both traveltime tomography and FWI, we advocate the following hybrid objective function, which the combines objective function for traveltime dominant FWI in equation 13 and conventional FWI in equation 1:

$$E(\mathbf{m}) = \frac{1-c}{2}\|\mathbf{u}(\mathbf{x},t)\,\mathbf{d}(\mathbf{x},t)\|^2 \qquad (16)$$
$$+ \frac{c}{2}\|\mathbf{d}(\mathbf{x},t) - \mathbf{d}_0(\mathbf{x},t)\|^2 \ ,$$

where $c = e^{-\mathbf{u}^2/\sigma^2}$ and σ is a constant that controls the width of this Gaussian. The adjoint source then becomes

$$\mathcal{A}(\mathbf{m}) \approx (1-c)\,\mathbf{u}^2 \mathbf{d} + c\,(\mathbf{d} - \mathbf{d}_0) \ , \qquad (17)$$

where we employ the same assumption as in equation 15.

When the traveltime misfit $\mathbf{u}$ is large, $c \to 0$ and hybrid FWI in equation 16 becomes traveltime dominate FWI in equation 13; when $\mathbf{u}$ is small, $c \to 1$ and hybrid FWI becomes conventional FWI in equation 1. Therefore, the inverse problem posed in equation 16 changes smoothly from traveltime dominant inversion to conventional FWI.

4 EXAMPLES

We test FWI with dynamic image warping for two different synthetic models: one with velocity anomalies and another with a deep reflector.


(a)

(b)

(c)
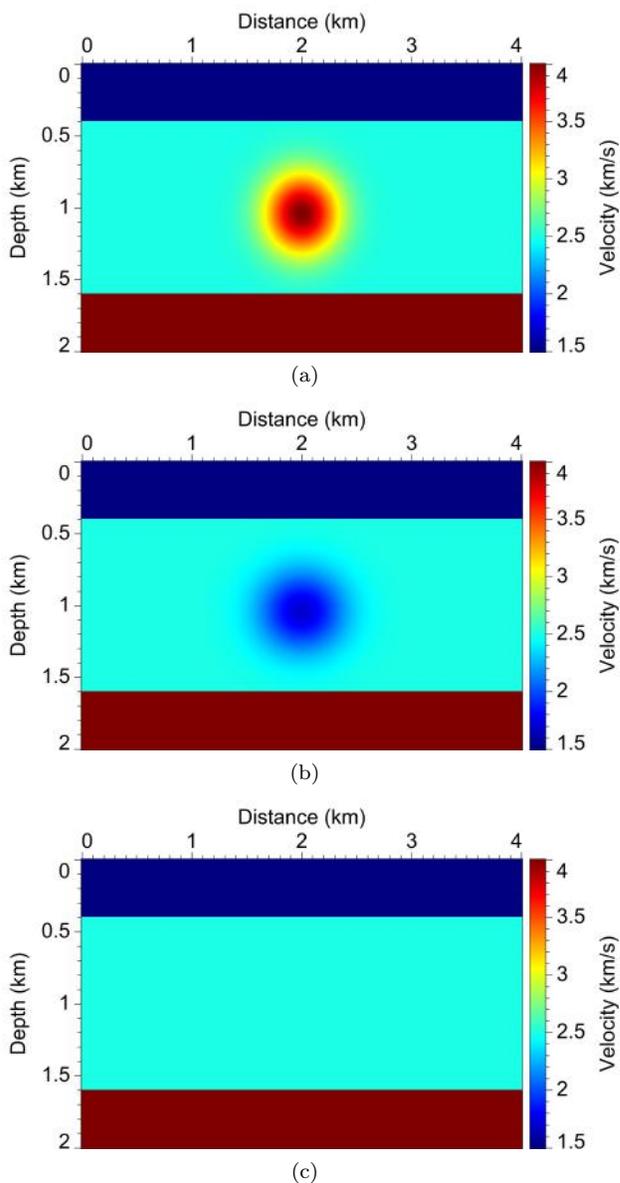
Figure 3. True velocity models that contain a high- (a) and a low-velocity (b) Gaussian anomalies. (c) Shows an initial model for inversion.

4.1 Example 1: velocity anomaly

Figures 3a and 3b display true velocity models that contain a high- or a low-velocity anomaly, respectively. The three-layer model without the anomalies shown in Figure 3c is the initial model for testing FWI. In this example, we use 24 shots uniformly distributed on the surface, and 8 Hz Ricker wavelet as the source for simulating wavefields.

Figure 4a displays a shot gather simulated for the model with a high-velocity anomaly, and Figure 4c shows the corresponding shot gather simulated in the

(a)

(b)

(c)
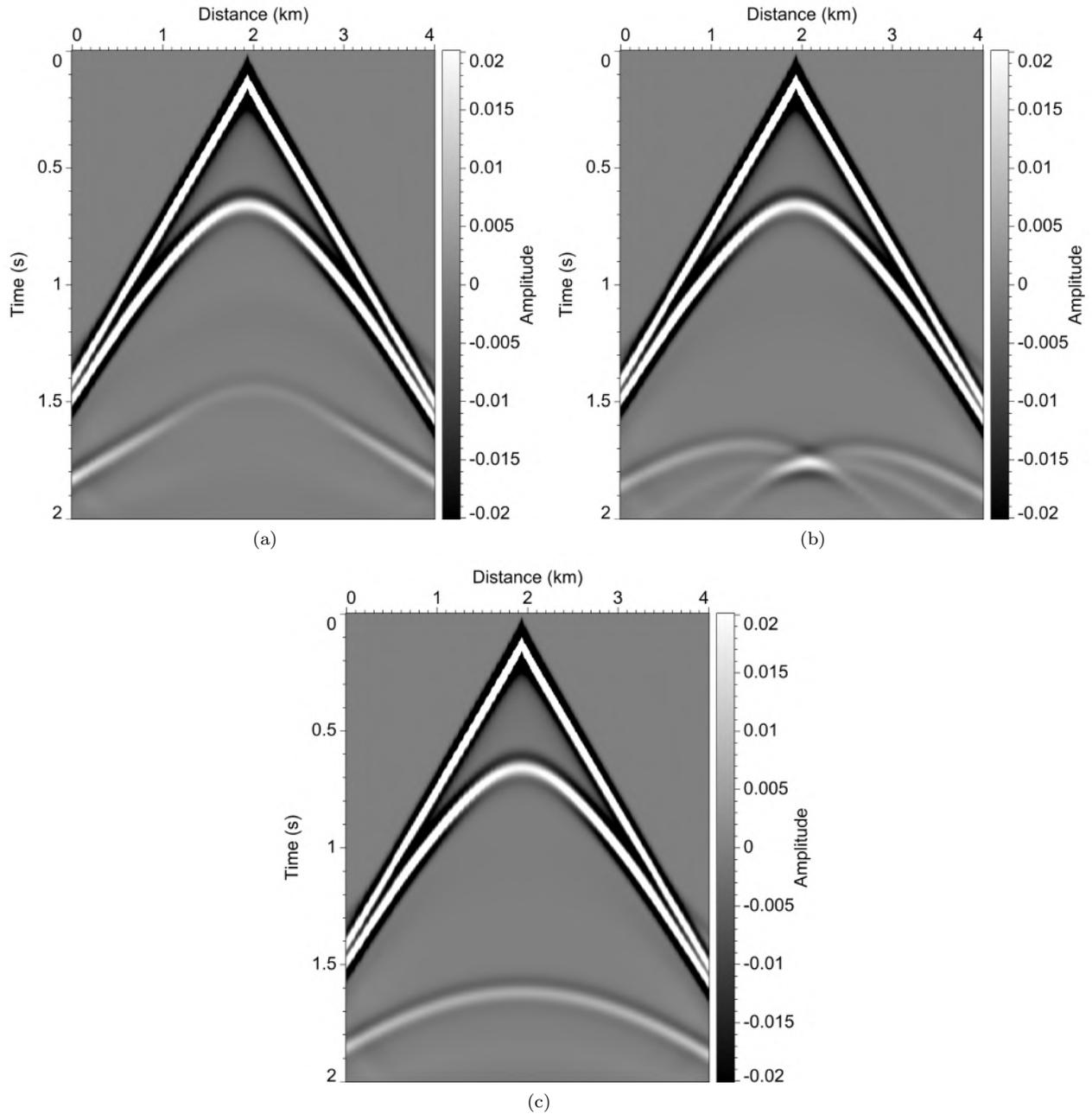
Figure 4. Recorded data $\mathbf{d}_0$ simulated in the true models with a high-velocity anomaly (a) and a low-velocity anomaly (b). Synthetic data $\mathbf{d}$ (c) is computed for the initial velocity model with no anomaly.

initial velocity model. The difference between Figures 4a and 4c is illustrated in Figure 5a. Similarly, Figure 4b displays a shot gather simulated in the model with a low-velocity anomaly, and Figure 5b illustrate the difference between Figures 4b and 4c.

The data residuals in Figure 5 indicate that cycle skipping occurs for both the high- and low-velocity anomalies. More quantitatively compelling evidence of

cycle skipping is the traveltime misfit $\mathbf{u}\,(\mathbf{x}, t)$ between the recorded data $\mathbf{d}_0$ and the synthetic data $\mathbf{d}$. Dynamic image warping is used to measure these time shifts. Figure 6a displays the absolute time shift caused by the high-velocity anomaly; Figure 6b shows the time shift normalized by the dominant period in the data. In Figure 6b we can observe shifts that are greater than one cycle, indicating the presence of cycle skipping. Like-
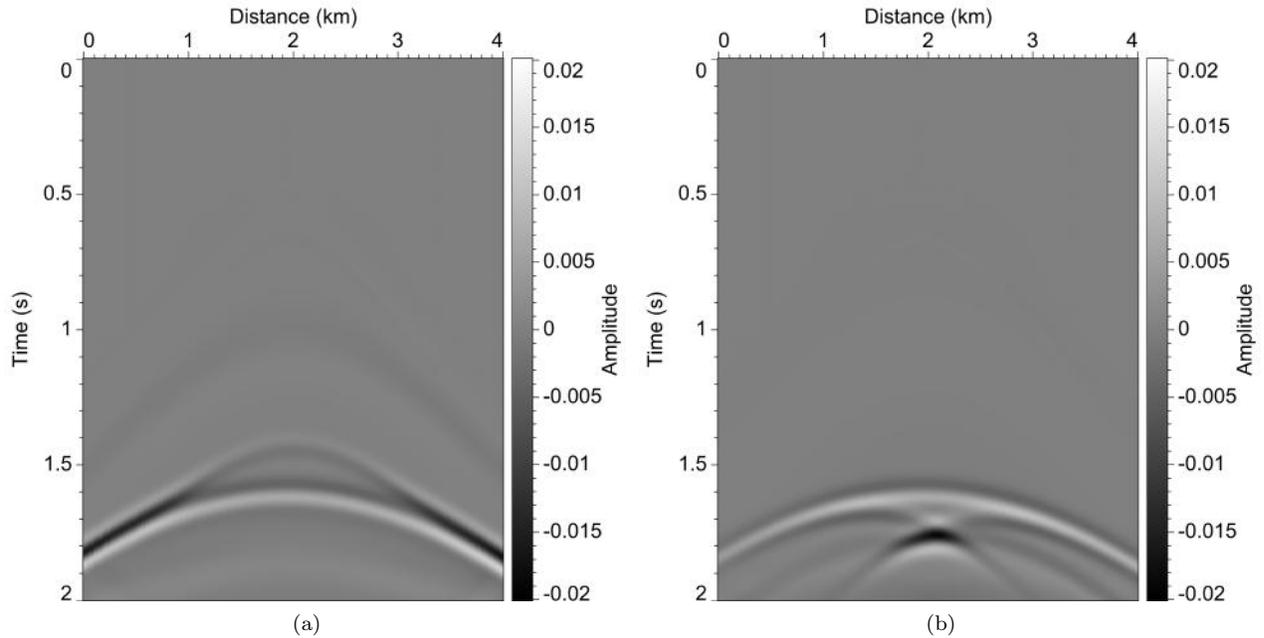
Figure 5. Data residuals $\mathbf{d} - \mathbf{d}_0$ corresponding to the models with high-velocity anomaly (a) and low-velocity anomaly (b).
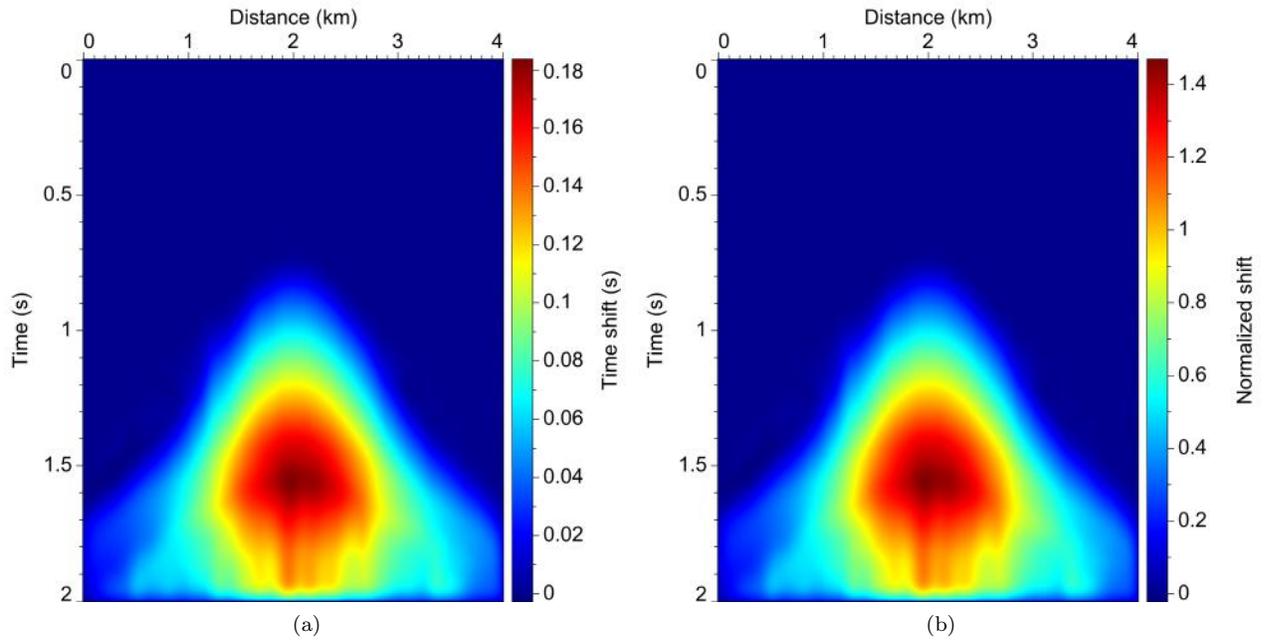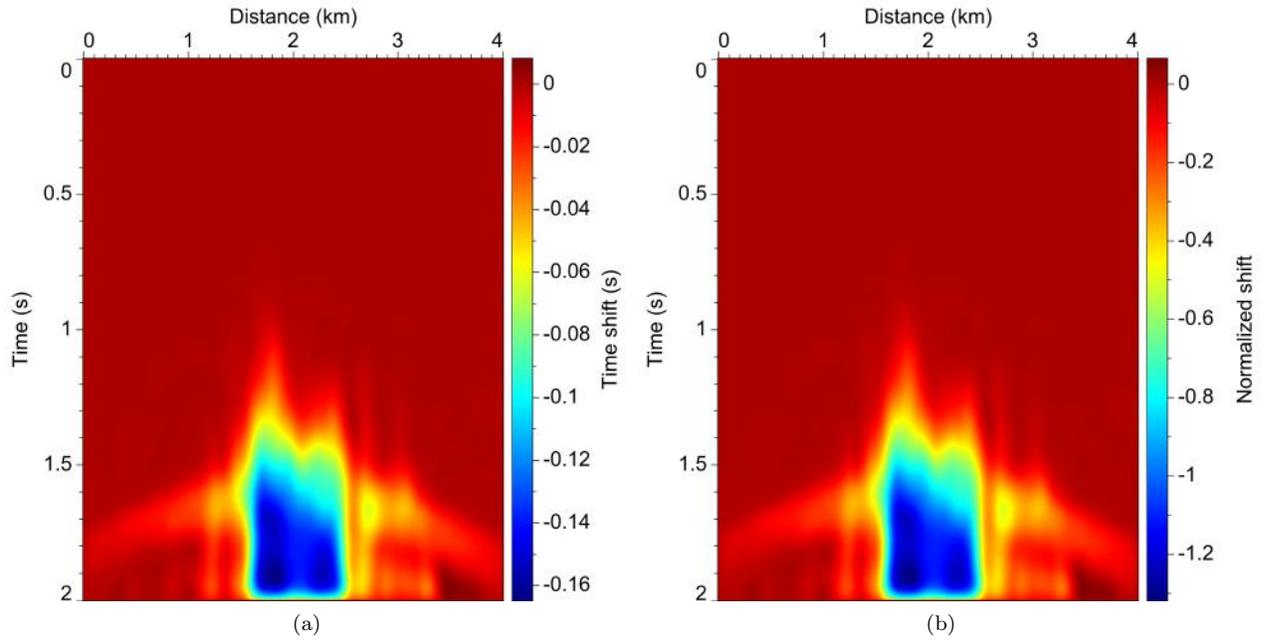

Figure 6. Traveltime misfit due to the high-velocity anomaly: absolute value (a) and shift (b) normalized by dominant period.

wise, Figure 7 shows the time shifts caused by the low-velocity anomaly and illustrates that cycle skipping occurs in the low-velocity anomaly as well.

Because of cycle skipping in this example, conventional FWI fails to recover neither of these velocity anomalies. Figures 8a and 9a show the conventional FWI velocity estimations for the high- and low-velocity anomalies, respectively. Comparing these estimates to the true velocities in Figures 3a and 3b, we observed that conventional FWI is trapped in local minima and updates the velocity in the direction that is opposite to the correct one.
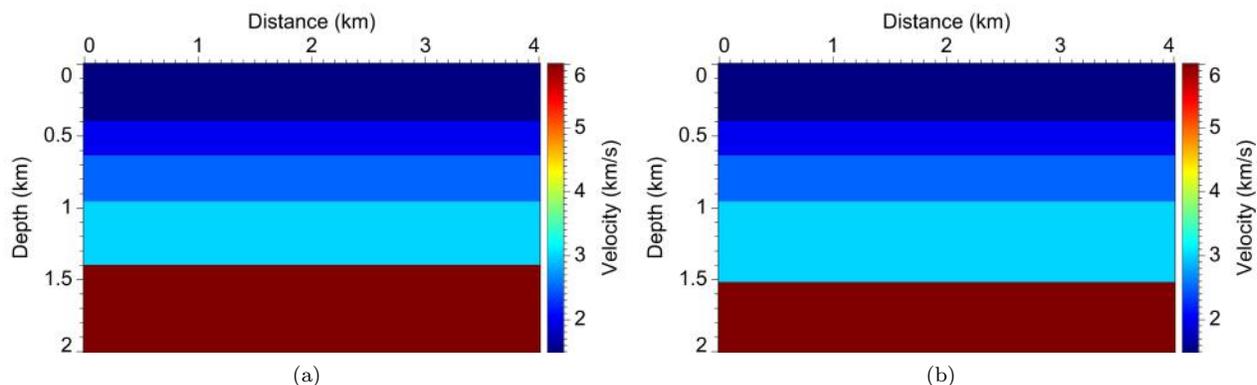
Figure 7. Traveltime misfit due to the low-velocity anomaly: absolute value (a) and shift (b) normalized by dominant period.
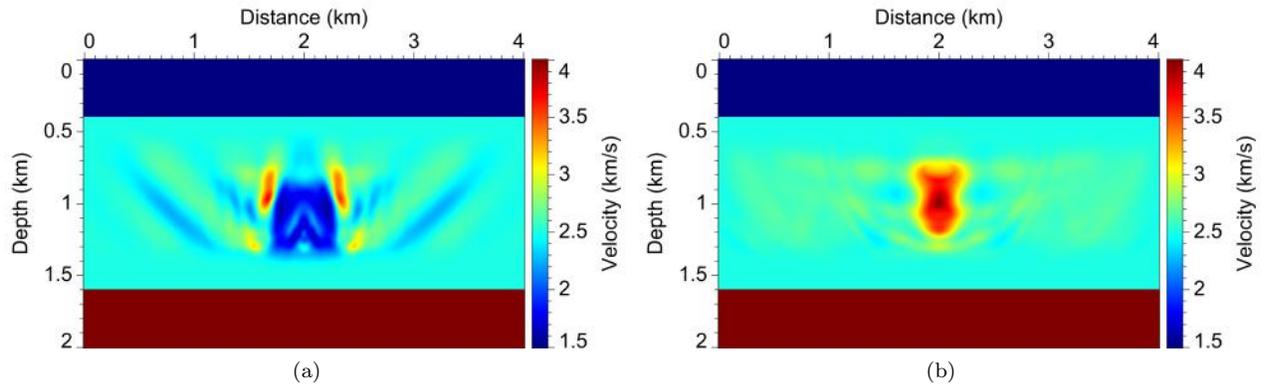
Figure 8. High-velocity anomaly with conventional FWI (a) and traveltime dominant FWI (b).
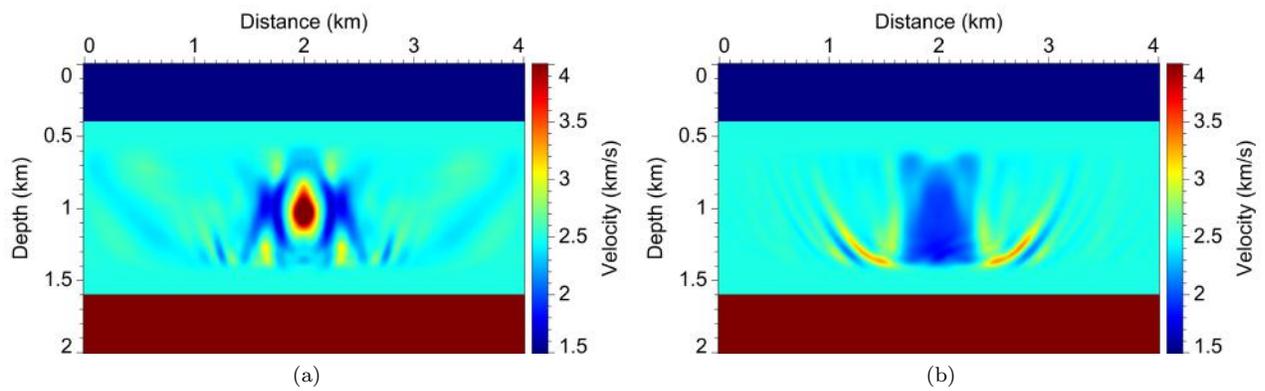
Figure 9. Low-velocity anomaly estimated with conventional FWI (a) and traveltime dominant FWI (b).

Figure 10. True velocity model (a) and initial velocity model (b). In the initial velocity model, the deepest reflector is misplaced by about 125 m.

Unlike conventional FWI, traveltime dominant FWI (equation 13) does not compute directly the data difference $\mathbf{d} - \mathbf{d}_0$; instead, it aims to minimize the traveltime misfit $\mathbf{u}(\mathbf{x}, t)$ between $\mathbf{d}_0$ and $\mathbf{d}$. The optimization problem implied by equation 13 avoids the cycle skipping by emphasizing the traveltime misfit $\mathbf{u}(\mathbf{x}, t)$ in the objective function.

For the high- and low-velocity anomalies, Figures 8b and 9b show respectively the updated models estimated by traveltime dominant FWI. Comparing these estimates with the true velocities shown in Figure 3a and 3b, we see that traveltime dominant FWI updates the velocities in the correct direction and recovers the anomalies reasonably well.

4.2 Example 2: deep reflector

Figure 10a displays a layered velocity model that is the true model in this second example. Figure 10b shows an initial velocity model, and the difference between the true and the initial models is in the depth of the deep reflector, which is misplaced by approximately 125 m. In this example we hope to recover the correct depth of the deep reflector from surface-recorded seismic data. We use 24 shots uniformly distributed on the surface, and a 15 Hz Ricker wavelet is used as the source.

Figures 11a and 11b display shot gathers simulated for the true and initial velocity models, respectively. The difference between the recorded data $\mathbf{d}_0$ (Figure 11a) and the synthetic data $\mathbf{d}$ (Figure 11b) is the initial data residual $\mathbf{d} - \mathbf{d}_0$, shown in Figure 12a. Two distinct events appear in the data residual, suggesting cycle skipping caused by the misplacement of the deep reflector. This cycle skipping can also be observed in the traveltime misfit $\mathbf{u}(\mathbf{x}, t)$ between the recorded data and the synthetic data. Figure 13a shows the traveltime misfit measured by dynamic image warping; Figure 13b plots the normalized shift, indicating the presence of cycle skip-

ping as the maximum normalized shift is greater than one cycle.

Due to cycle skipping, conventional FWI fails to correct the misplaced deep interface, as illustrated in Figure 14a. We computed the synthetic data $\mathbf{d}$ with the updated velocity model (Figure 14a) and show the data residual $\mathbf{d} - \mathbf{d}_0$ in Figure 12b. Compared to the initial residual (Figure 12a), the new residual (Figure 12b) shows smaller magnitude; however, the two events that are caused by cycle skipping become separated farther, indicating that conventional FWI updates the model in a wrong direction.

Figure 14b shows the velocity model updated by traveltime dominant FWI (equation 13) in 10 iterations. This updated model correctly positions the deep reflector while showing a smooth interface. In the updated data residual shown in Figure 12c, we can find only one event, suggesting that the inversion avoids cycle skipping. However, this new residual still contains high amplitudes due to the smooth update. Although equation 13 involves both the traveltime misfit $\mathbf{u}$ and the synthetic data $\mathbf{d}$, it is the misfit that most significantly controls the objective function. In other words, traveltime dominant FWI uses primarily the traveltime information and therefore performs like a traveltime tomography. This explains why traveltime dominant FWI cannot recover the sharp interface.

We can improve the inversion implied by equation 13 by applying a hybrid FWI formulated in equation 16. Figure 14c displays the velocity model estimated by the hybrid FWI in 10 iterations. Compared with Figures 14a and 14b, Figure 14c not only places the deep reflector at the correct depth, but also recovers the sharp boundary. In early iterations, the hybrid FWI behaves predominantly like traveltime dominant FWI due to a large traveltime misfit; in this stage, the hybrid FWI mainly moves the deep reflector towards the correct depth. As the traveltime misfit decreases, the hybrid FWI gradually becomes conventional FWI;
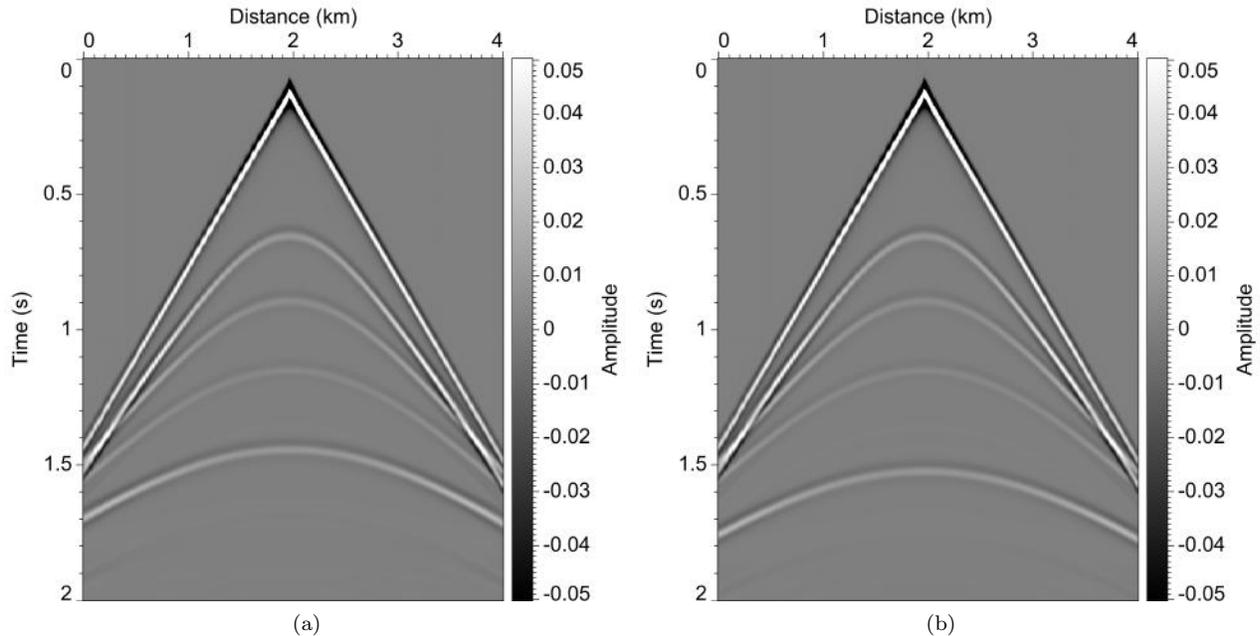
Figure 11. Recorded data (a) simulated for the true velocity model (Figure 10a) and synthetic data (b) computed for the initial velocity model (Figure 10b).

in this later stage, the hybrid FWI sharpens the reflecting interface. Figure 12d shows the data residual $\mathbf{d} - \mathbf{d}_0$ corresponding to the hybrid FWI model. The data residual in Figure 12d shows a substantially lower magnitude that is apparent in the other three data residuals.

5 CONCLUSION

We have proposed using DIW as a tool to estimate traveltime misfit between recorded seismic data and synthetic data in FWI. As a global optimization, DIW overcomes the cycle-skipping problem in common traveltime estimations based on local crosscorrelations. We use the DIW-derived traveltime misfit in FWI by defining two objective functions that involve the misfit. Traveltime dominant FWI implied by equation 13 can avoid cycle skipping when the initial model is far from the true one; however, it does not provide high-wavenumber details because it predominantly uses traveltime information in inversion. The hybrid FWI implied by equation 16 combines the benefits of both traveltime dominant FWI and conventional FWI.

The adjoint-state method is used to compute the gradient $\frac{\partial E(\mathbf{m})}{\partial \mathbf{m}}$. In order to conveniently compute adjoint sources, we made an assumption that $\frac{\partial \mathbf{u}}{\partial \mathbf{d}}$ is negligible. This assumption requires further investigation. However, the alternative but useful objective functions may exist with assumptions more valid than those proposed in this paper.

6 ACKNOWLEDGMENTS

REFERENCES

Crase, E., A. Pica, M. Noble, J. McDonald, and A. Tarantola, 1990, Robust elastic nonlinear waveform inversion: Application to real data: Geophysics, **55**, 527–538.

de Hoop, M. V., and R. D. van der Hilst, 2005, On sensitivity kernels for 'wave-equation' transmission tomography: Geophysical Journal International, **160**, 621–633.

Guitton, A., and W. W. Symes, 2003, Robust inversion of seismic data using the huber norm: Geophysics, **68**, 1310–1319.

Hale, D., 2012, Dynamic warping of seismic images: CWP Report, **723**.

Loris, I., G. Nolet, I. Daubechies, and F. A. Dahlen, 2007, Tomographic inversion using L1-norm regularization of wavelet coefficients: Geophysical Journal International, **170**, 359–370.

Luo, Y., and G. T. Schuster, 1991, Wave-equation traveltime inversion: Geophysics, **56**, 645–653.

Pratt, R. G., 1999, Seismic waveform inversion in the
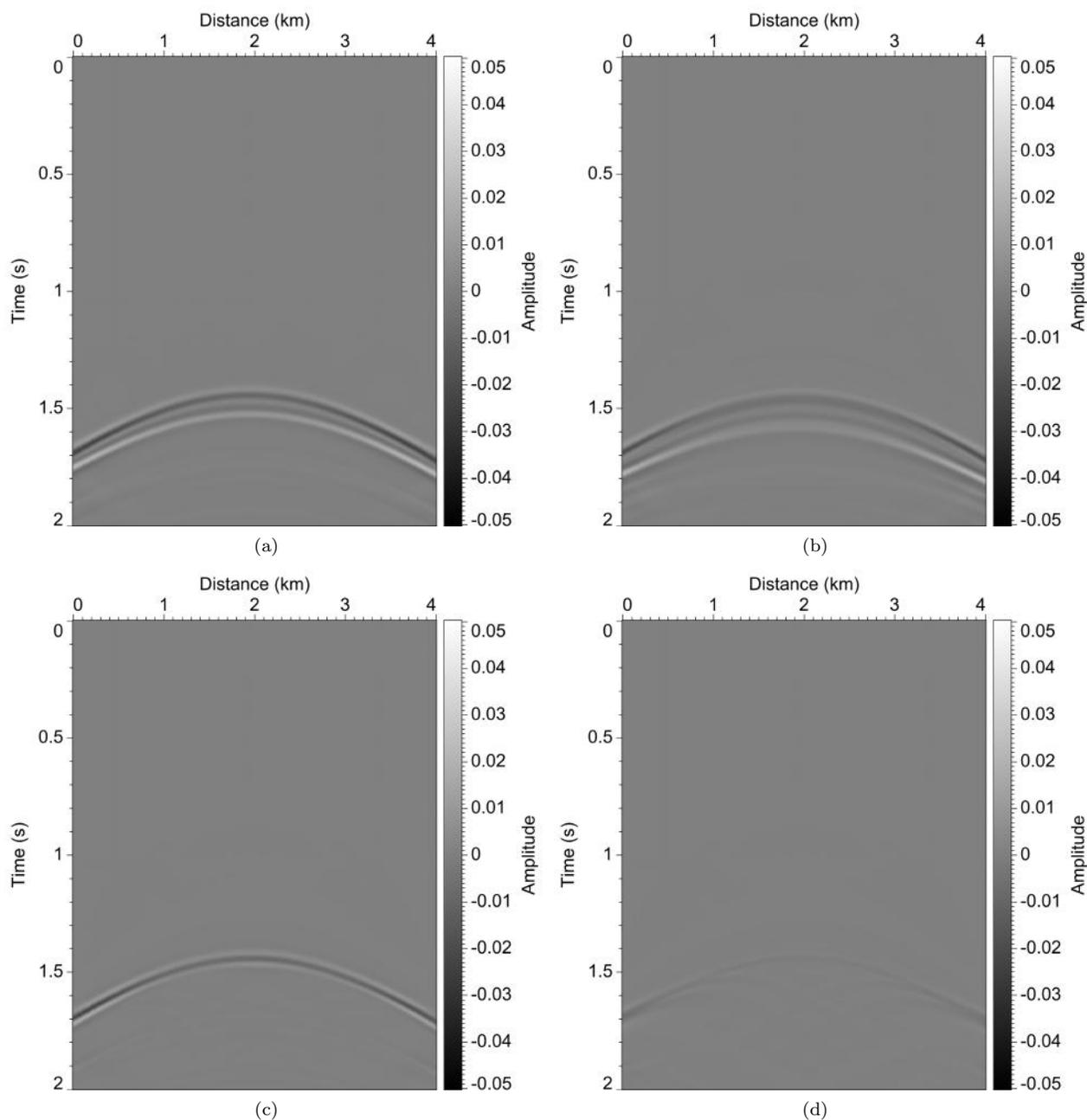
Figure 12. Data residuals $\mathbf{d} - \mathbf{d}_0$ computed for the initial velocity model (a), and for velocity models updated by the conventional FWI (b), traveltime dominant FWI (c), and the hybrid FWI (d).

frequency domain, part 1: Theory and verification in a physical scale model: Geophysics, **64**, 888.

Sakoe, H., and S. Chiba, 1978, Dynamic programming algorithm optimization for spoken word recognition: IEEE Transactions on Acoustics, Speech, and Signal Processing, **26**, 43–49.

Snieder, R., M. Y. Xie, A. Pica, and A. Tarantola, 1989, Retrieving both the impedance contrast and background velocity: A global strategy for the seismic reflection problem: Geophysics, **54**, 991–1000.
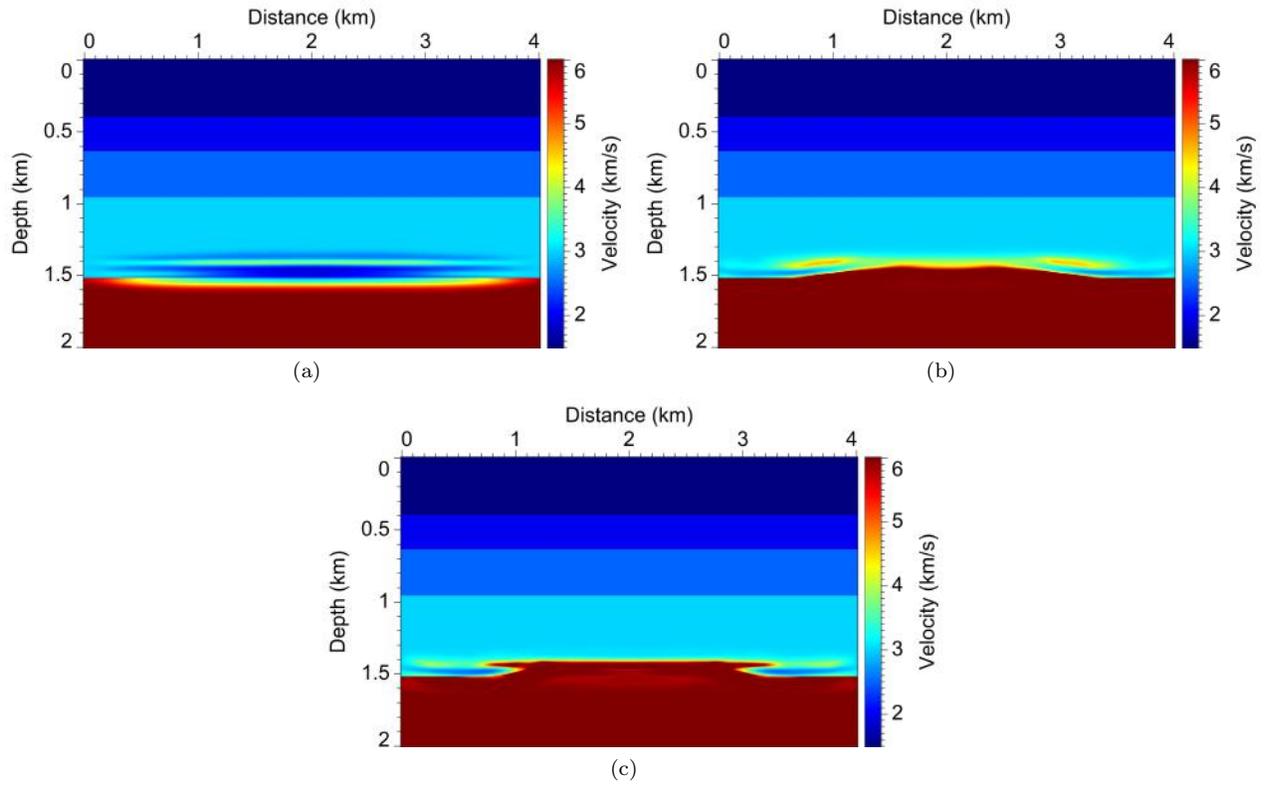
Stork, C., 1992, Reflection tomography in the postmigrated domain: Geophysics, **57**, 680–692.

Tarantola, A., 1984, Inversion of seismic-reflection data in the acoustic approximation: Geophysics, **49**, 1259–1266.

———, 2005, Inverse problem theory and methods for
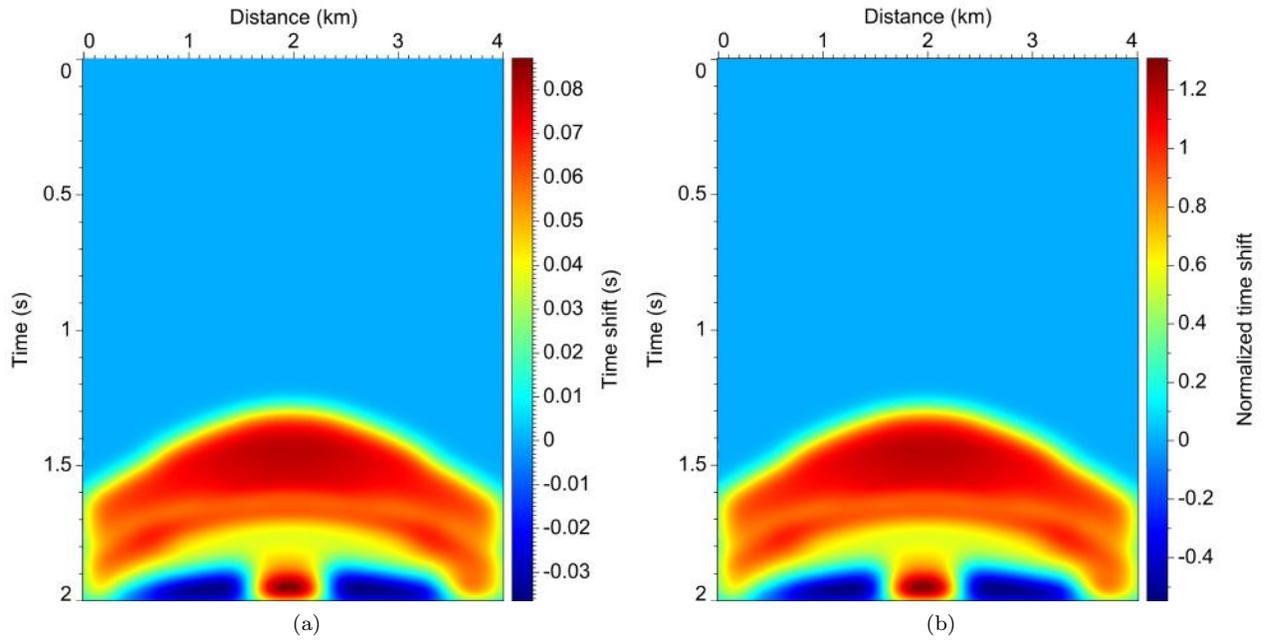
Figure 13. Traveltime misfit detected by DIW: absolute shift (a) and shift (b) normalized by dominant period.

Figure 14. Velocities estimated by conventional FWI (a), traveltime dominant FWI (b), and the hybrid FWI (c).

model parameter estimation: Society for Industrial and Applied Mathematics.

Tromp, J., C. Tape, and Q. Liu, 2005, Seismic tomography, adjoint methods, time reversal and banana-doughnut kernels: Geophysical Journal International, **160**, 195–216.

van Leeuwen, T., and W. A. Mulder, 2010, A correlation-based misift criterion for wave-equation traveltime tomography: Geophysical Journal International, **182**, 1383–1394.

Vasco, D., and E. Majer, 1993, Wavepath travel-time tomography: Geophysical Journal International, **115**, 1055–1069.

Vigh, D., and E. Starr, 2008, 3D prestack plane-wave, full-waveform inversion: Geophysics, **73**, VE135–VE144.

Woodward, M. J., 1992, Wave-equation tomography: Geophysics, **57**, 15–26.

Zelt, C. A., and P. J. Barton, 1998, Three-dimensional seismic refraction tomography: A comparison of two methods applied to data from the faeroe basin: J. Geophys. Res., **103**, 7187–7210.